HGO-CNN: HYBRID GENERIC-ORGAN CONVOLUTIONAL NEURAL NETWORK FOR MULTI-ORGAN PLANT CLASSIFICATION

Sue Han Lee[⋆] *Yang Loong Chang*[⋆] *Chee Seng Chan*[⋆] *Paolo Remagnino*[‡]

[⋆]Centre of Image & Signal Processing, Fac. Comp. Sci. & Info. Tech., University of Malaya, Malaysia
[‡]The Robot Vision Team, Comp. Sci. Dept., Kingston University, United Kingdom
{*leesuehan,yangloong@siswa.um.edu.my; cs.chan@um.edu.my; p.remagnino@kingston.ac.uk*}

ABSTRACT

Classification of plants based on a multi-organ approach is very challenging. Although additional data provides more information that might help to disambiguate between species, the variability in shape and appearance in plant organs also raises the degree of complexity of the problem. Existing approaches focus mainly on generic features for species classification, disregarding the features representing the organs. In fact, plants are complex entities sustained by a number of organ systems. In our approach, we exploit the PlantClef2015 benchmark, and introduce a hybrid generic-organ convolutional neural network (HGO-CNN), which takes into account both organ and generic information, combining them using a new feature fusion scheme for species classification. We show that our proposed method outperforms the state-of-the-art results.

Index Terms— Plant classification, deep learning

1. INTRODUCTION

Botanists classify plant species by observing plant organs: the stem, flowers, fruits and leaves of the studied plant. Among all organs, the leaf and their characters are studied extensively [1, 2]. Computer based methods have been designed to support botanists [3–8]. Existing literature is concerned with plant identification using automated pattern analysis based on leaf characters. Although the structural features of a leaf are important in the plant identification task, for certain plants, such as deciduous or semi-evergreen plants, leaves are not visible or available over different periods of the years. In these cases, multiple organs are required to identify the correct species. In 2013, the LifeClef challenge [9] provided the first multi-organ plant dataset. This was the first multi-organ plant classification benchmark in computer vision.

However, it is a challenging task to classify plants based on a multi-organ approach. For example, in Fig 1, we can observe the large variability in the appearance of plant organs. Even within the same organ, large differences can occur. Furthermore, for images taken in the outdoor field, the clutter in the background makes more difficult recognizing plant species.

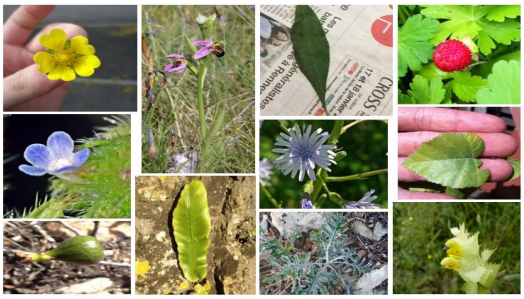

Fig. 1: Large variability in the appearance of plant organs

Deep learning has shown a notable success in large-scale image recognition [10, 11]. In the multi-organ plant identification task, many of the existing methods [12, 13] employ deep learning to train an N-class species classifier, irrespective of the organ or organ structure. The features learned based on this approach tend to be generic. In [12], the training of a generic network using a deeper learning network, the GoogLenet showed the best result for the Plantclef 2015 dataset [9]. Although generic features can model target species classes, they might not be able to provide an appropriate description for a plant. For example, for a leaf image taken with a noisy background with text, such as the leaf on the newspaper shown in Fig. 1, generic features focus on the holistic representation of the image. In such case, text might be considered erroneously as one of the discriminative features for the species. This is not surprising, as a generic network learns irrelevant features, especially when they appear to be discriminative among species. Hence, a novel approach is required that can go beyond the generic description of the plant and provide a better reasoning to model plant species.

Generally, botanists can classify plants by observing and studying their features, usually using all the plant organs. Plant organs are known prior to explore the characteristic of a species. For instant, when botanists study a **leaf**, they focus on the leaf characters such as its *margin* or *venation* patterns, and, when they study a **flower**, they focus on the characteris-

* Sue Han Lee and Yang Loong Chang contributed equally to this paper.

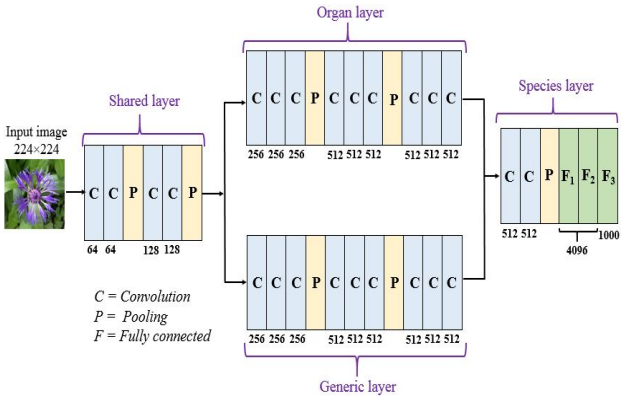

Fig. 2: The architecture of the proposed HGO-CNN.

tics of its *petals*, *sepals* and *stamen* to identify unknown plant species. So, it is logical to believe that a better recognition method for plant species might require prior information of their organs.

In this paper, we propose a novel architecture, and call it a hybrid generic-organ convolutional neural network, abbreviated HGO-CNN. HGO-CNN integrates both generic and organ-based information for the multi-organ plant classification task. Specifically, it is able to encapsulate organ and generic information prior to species inference for multi-organ plant classification. This paper has two main contributions. First, HGO-CNN introduces a novel classification model. It extracts prior organ information, and, classifies images of plants based on the correlation between the chosen organ and generic-based features. Second, HGO-CNN introduces a new fusion scheme, to learn the correlation between an organ and generic components. We show that both components can be independently pretrained and integrated to form one single architecture with the whole feature extraction and species classification operations jointly trained end-to-end.

2. THE HGO-CNN

The proposed HGO-CNN comprises four layers or components: (i) a shared layer, (ii) an organ layer, (iii) a generic layer, and (iv) a species layer. The rationale behind proposing a shared layer is inspired by: (1) the work of [14, 15], who demonstrated that bottom layers in deep networks respond to low-level features, such as corners and edges, in turn crucial to the classification of any high level features, and, (2) the fact that such layers help reducing the number of training parameters. Input to our HGO-CNN is a color image of 224 $\times$ 224 pixels. For the convolutional layer, we utilise 3 $\times$ 3 convolution filters with spatial resolution preserved using stride 1. Max pooling is performed using a 2 $\times$ 2 pixel window with stride 2. Three fully connected layers, which have 4096, 4096 and 1000 channels respectively, follow behind the stacks of convolutional layers. The final layer is the softmax layer.

2.1. Implementation Details

In order to train the HGO-CNN to capture prior organ information for species classification, we propose a feature fusion scheme. It is based on a novel step-by-step training strategy. The HGO-CNN is trained using the following steps:

Pre-Training CNN layers HGO-CNN uses a two path CNN for the purpose of training generic and organ based features in a later stage. This two path CNN is similar to the architecture depicted in Fig. 2, except that, it does not include the interconnection between paths, and, each path has its own fully connected layers. These are initially pre-trained using the ImageNet challenge dataset [16].

Organ layer After pre-trained, one of the CNN path is repurposed to train on the organ task. This organ layer is trained together with the shared layer, using seven kinds of predefined organ labels. We obtain organ-based feature maps $\mathbf{x_{org}}$ in the $H \times W \times Z$ 3D cuboid, where H, W and Z are the width, height and number of channels of the respective feature maps.

Generic & species layer After training the organ layer, we train the species layer based on the species labels. Here, we encapsulate both organ and generic information prior to species classification. We train the species layer based on the correspondence of these two components – organ and generic. During the species layer training, the generic layer is fine-tuned using a lower learning rate, and, output a generic feature maps $\mathbf{x_{gen}}$ in the $H^* \times W^* \times Z^*$ 3D cuboid. To allow both organ and generic layers to share the common proceeding layers, we keep the shared layer's weights to be consistent. This is achieved by setting their learning rate to zero. To put in correspondence between both organ and generic components, a fusion function $g : \mathbf{x_{org}}, \mathbf{x_{gen}} \rightarrow \mathbf{y}$ at layer L is learned to produced organ and a generic correlation feature maps $\mathbf{y}$ in the $H^{"} \times W^{"} \times Z^{"}$ 3D cuboid. In our model, L is the last convolutional layer for both components. Since, we fuse both components in the same layer L, where both feature maps have the same dimension, $h = H = H^*$, $w = W = W^*$ and $z = Z = Z^*$. In the species layer, g firstly concatenates these two sets of feature maps along the channel axis, forming a stacked data $\mathbf{x_{cat}} = [\mathbf{x_{gen}}, \mathbf{x_{org}}]$ in the $h \times w \times 2z$ 3D cuboid. Then, $\mathbf{x_{cat}}$ will subsequently convolves with a set of filters $\mathbf{f}$ and biases $\mathbf{b}$.

$$\mathbf{y} = \mathbf{x_{cat}} * \mathbf{f} + \mathbf{b} \tag{1}$$

$\mathbf{f}$ is a filter bank of size N, and, each filter is in the $p \times q \times 2z$ 3D cuboid. Size of $\mathbf{b}$ is equal to the number of filters. In our model, we set $N = z$ so that we can reduce the dimensionality of the output feature maps, while, at the same time, modeling the correspondence between the two feature maps $\mathbf{x_{gen}}$ and $\mathbf{x_{org}}$. We set the learning rate of the new randomly-initialised species layers to 10 times higher than the preceding initialised

layers and fix the weight of the organ layer when optimizing the model with respect to species classes.

3. DATASETS AND EVALUATION METRICS

Dataset. The PlantClef2015 dataset has 1000 plant species classes. Training and testing data comprises 91759 and 21446 images respectively. Each image is associated with single organ type (branch, entire, flower, fruit, leaf, stem or leaf scan).

Evaluation metrics. Two evaluation metrics are employed: the *image-centered* and the *observation* score [9]. The purpose of the observation score is to evaluate the ability of a model predicting correct species labels to all the users. Observation score calculates the mean of the average classification rate per user as defined:

$$S_{obs} = \frac{1}{U}\sum_{u=1}^{U}\frac{1}{P_u}\sum_{p=1}^{P_u}S_{u,p} \tag{2}$$

where U: number of users, P_u: number of individual plants observed by the u-th user, $S_{u,p}$: score between 0 and 1 equals to the inverse of the rank of the correct species (for the p-th plant observed by the u-th user). Each query observation is composed of multiple images. To compute $S_{u,p}$, we adopt the Borda count (BD) and the majority voting (MAV) based approaches to combine the scores of multiple images:

$$BD = \frac{1}{n}\sum_{k=1}^{n} score_k \tag{3}$$

$$MAV = \max_{1\leq k\leq n} score_k \tag{4}$$

where n: total images per query observation. $score$: softmax output score which describes the ranking of the species.

Next, image-centered score evaluates the ability of a system providing the correct species labels based on a single plant observation. It calculates the average classification rate on each individual plant as defined:

$$S_{img} = \frac{1}{U}\sum_{u=1}^{U}\frac{1}{P_u}\sum_{p=1}^{P_u}\frac{1}{N_{u,p}}\sum_{n=1}^{N_{u,p}}S_{u,p,n} \tag{5}$$

where U and P_u are explained above. $N_{u,p}$ is the number of pictures taken from the p-th plant observed by the u-th user, $S_{u,p,n}$ is the score between 0 and 1 equals to the inverse of the rank of the correct species (for the n-th picture taken from the p-th plant observed by the u-th user). We compute the rank of the correct species based on its softmax scores.

4. EXPERIMENTS

We train our model using the *Caffe* [17] framework. For the parameter setting in training, we employ fixed learning policy.

Table 1: Performance comparison with other proposed methods. Note that, M-S = Multi-scale.

Method	S_{obs}	S_{img}
GoogLeNet + Fisher Vectors (BD) [13]	0.592	-
GoogLeNet (MAV) [13]	0.609	0.581
GoogLeNet (content+ domain) [19]	0.633	-
GoogLeNet + softmax normalization [19]	0.624	0.590
5-fold GoogLeNet (MAV) [12]	0.667	0.652
5-fold GoogLeNet (BD) [12]	0.663	0.652
VGG-16 net(MAV)	0.663	0.638
VGG-16 net(BD)	0.664	0.638
HGO-CNN(MAV)	0.671	0.647
HGO-CNN(BD)	0.673	0.647
M-S HGO-CNN(MAV)	**0.715**	**0.690**
M-S HGO-CNN(BD)	**0.717**	**0.690**

We set the learning rate to 0.01, and then decrease it by a factor of 10 when the validation set accuracy stop improving. The momentum is set to 0.9 and weight decay to 0.0001. All the networks are trained by back propagation using stochastic gradient descent [18]. We improve the generalization of the model by randomly cropping and mirroring the input image during training.

4.1. Performance Evaluation

We compare our HGO-CNN with the current state-of-the-art (SOTA) methods [12, 13, 19]. We also compare with the VGG-16 net [20], which is fine tuned and trained purely on species labels using the PlantClef2015 dataset. This is to measure the contribution of correlation between organ and generic components in the plant species classification. Table 1 shows the comparison results. We observe that the HGO-CNN model achieves a higher score compared to the VGG-16 net. This confirms the importance of organ features used to discriminate between plant species compared to using solely generic information for plant classification.

To increase the robustness of the system in recognising multi-organ plant images, a multi-scale training is adopted. We isotropically rescale the training images into three different sizes: 256, 384 and 512. Then, for 384 and 512 image sizes, we crop 256×256 center pixels. During network training, 224×224 pixels are randomly cropped from the rescaled images and fed into the network. We call this network a multi-scale HGO-CNN (M-S HGO-CNN). During the class prediction phase, we do apply a similar multi-scale process to obtain three sets of testing images for a query image. An averaging fusion method is then used to combine their softmax scores to output a final result for a query image. It is noticeable that our M-S HGO-CNN model outperforms all the SOTA methods, achieving the best results for the PlantClef2015 dataset.

4.2. Detailed Scores of Each Plant Organ

In this section, we analyse the classification performance of each organ based on the image-centered score, S_{img}. Table

Table 2: Classification performance comparison of each contents based on S_{img}.

Method	Branch	Entire	Flower	Fruit	Leaf	LeafScan	Stem
Choi [12]	0.498	0.531	0.784	0.602	0.600	0.766	0.326
Ge et al. [19]	0.416	0.448	0.738	0.558	0.524	0.694	0.291
Champ et al. [13]	0.398	0.453	0.723	0.559	0.501	0.713	0.302
Le et al. [21]	0.051	0.084	0.207	0.125	0.342	0.737	0.164
VGG-16 net	0.491	0.522	0.777	0.585	0.591	0.747	0.337
HGO-CNN	0.522	0.532	0.779	0.604	0.607	0.690	0.326
M-S HGO-CNN	**0.568**	**0.603**	**0.798**	**0.653**	**0.652**	**0.803**	**0.411**

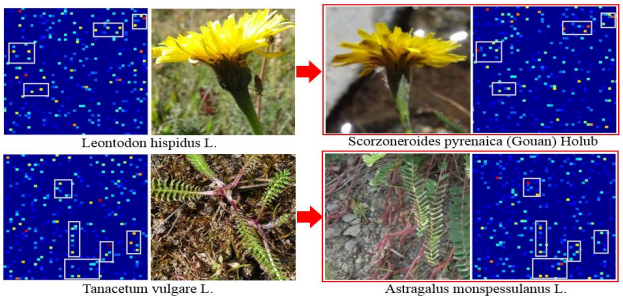

Fig. 3: Misclassified examples. The projected F_2 features of misclassified images (right) are found having almost similar feature patterns to the wrongly classified species classes (left).

2 illustrates the comparison results. We observe that both of our proposed model, HGO-CNN and M-S HGO-CNN show scanned-leaf and flower are the most effective organs compared to others for plant identification. This is similar to the results reported in [9]. Our HGO-CNN shows a higher identification score for 'Flower' category compared to 'LeafScan'. In addition, using multi-scale training, M-S HGO-CNN shows a major improvement in 'LeafScan' category. This indicates that multi-scale training data could further improve the feature representation for multi-organ plant images. In overall, our M-S HGO-CNN achieves the highest S_{img} compared to other SOTAs. Although M-S HGO-CNN leads to a better result for 'Stem', it is still considered as the least informative one compared to other organs. This might be due to the intra and interspecies diversity of plants in nature, resulting in stem not vivid enough for species inference.

4.3. Failure Analysis

We performed failure analysis and observed that these wrongly classified test images have very similar feature patterns with the training images from its wrongly classified species classes. For example, in Fig. 3, the *Leontodon hispidus L.* that was misclassified as *Scorzoneroides pyrenaica (Gouan) Holub* has very similar visual appearances at the parts of flower, particularly the color or shape of petals. Through feature visualisation of F_2 layer, some similar feature patterns (drawn in white bounding boxes) can be observed as well. However, this mistake is understandable as plants in nature have small interspecies variation, and, generally under such circumstances, more sophisticated plant morphology is used.

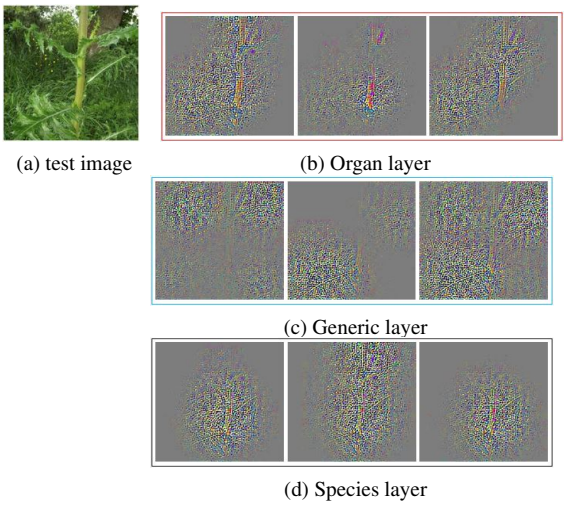

Fig. 4: Visualisation of the last convolution of generic, organ and species layer for the test image. Color contrast is digitally enhanced. Figure is best viewed in electronic form.

4.4. Qualitative Analysis

We visualise the characteristic of organ, generic and species layer based on the deconvolution approach [14]. We subsample the top 3 activation feature maps in each layer and reconstructed back to image pixels. Fig. 4 shows the deconvolution results. We observe that both organ and generic-based features show complementary information, in which organ layer is mainly focusing on the tree branch while generic layer stimulates at the twig. Based on the correlation strategy, the species layer encapsulates both information and reveals the portions that best represent the plant image.

5. CONCLUSION

We have presented HGO-CNN, a new approach that uses an end-to-end deep neural network to integrate both organ and generic features, and, capture the correlation of these complementary information for species classification. Experiments on the PlantClef 2015 benchmark show the robustness of HGO-CNN in multi-organ plant classification. It is worth noting that using multi-scale training can further boost up the discriminative power of the model. Based on our findings, it is clear that, not all sets of plant organs are useful for species inference, and as such, we will further our research to investigate a system that offers extra flexibility in learning the relationship between organs, targeting only discriminative types of organs that represent best for plant species.

6. ACKNOWLEDGEMENTS

This research is supported by the UM PPP Grant PG007-2016A, and the used Titan X was donated by NVIDIA.

7. REFERENCES

[1] James Clarke, Sarah Barman, Paolo Remagnino, Ken Bailey, Don Kirkup, Simon Mayo, and Paul Wilkin, "Venation pattern analysis of leaf images," in *Advances in Visual Computing*, pp. 427–436. Springer, 2006.

[2] Beth Ellis, Douglas Daly, Leo Hickey, Kirk Johnson, John Mitchell, Peter Wilf, and Scott Wing, *Manual of leaf architecture*, vol. 190, Cornell University Press Ithaca, 2009.

[3] Sue Han Lee, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in *ICIP*, 2015, pp. 452–456.

[4] Cem Kalyoncu and Önsen Toygar, "Geometric leaf classification," *Computer Vision and Image Understanding*, vol. 133, pp. 102–109, 2015.

[5] Paolo Remagnino, Simon Mayo, Paul Wilkin, James Cope, and Don Kirkup, *Computational Botany: Methods for Automated Species Identification*, Springer, 2017.

[6] David Hall, Chris McCool, Feras Dayoub, Niko Sunderhauf, and Ben Upcroft, "Evaluation of features for leaf classification in challenging conditions," in *WACV*. IEEE, 2015, pp. 797–804.

[7] Y.G. Naresh and H.S. Nagendraswamy, "Classification of medicinal plants: An approach using modified lbp with symbolic representation," *Neurocomputing*, vol. 173, pp. 1789–1797, 2016.

[8] Zhe Tang, Yuancheng Su, Meng Joo Er, Fang Qi, Li Zhang, and Jianyong Zhou, "A local binary pattern based texture descriptors for classification of tea leaves," *Neurocomputing*, vol. 168, pp. 1011–1023, 2015.

[9] Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Robert Planqué, Andreas Rauber, Simone Palazzo, Bob Fisher, et al., "Lifeclef 2015: multimedia life species identification challenges," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pp. 462–483. Springer, 2015.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[11] Peng Tang, Xinggang Wang, Bin Feng, and Wenyu Liu, "Learning multi-instance deep discriminative patterns for image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3385–3396, 2017.

[12] Sungbin Choi, "Plant identification with deep convolutional neural network: Snumedinfo at lifeclef plant identification task 2015," in *Working notes of CLEF 2015 conference*, 2015.

[13] Julien Champ, Titouan Lorieul, Maximilien Servajean, and Alexis Joly, "A comparative study of fine-grained classification methods in the context of the lifeclef plant identification challenge 2015," in *Working notes of CLEF 2015 conference*, 2015.

[14] Matthew Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks," in *ECCV 2014*, pp. 818–833. Springer, 2014.

[15] Zhicheng Yan, Hao Zhang, Robinson Piramuthu, Vignesh Jagadeesh, Dennis DeCoste, Wei Di, and Yizhou Yu, "Hd-cnn: hierarchical deep convolutional neural networks for large scale visual recognition," in *ICCV*, 2015, pp. 2740–2748.

[16] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[17] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *ACM-MM*, 2014, pp. 675–678.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[19] ZongYuan Ge, Chris Mccool, Conrad Sanderson, and Peter Corke, "Content specific feature learning for fine-grained plant classification," in *Working notes of CLEF 2015 conference*, 2015.

[20] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[21] Thi-Lan Le, DN Dng, Hai Vu, and Thanh-Nhan Nguyen, "Mica at lifeclef 2015: Multi-organ plant identification," in *Working notes of CLEF 2015 conference*, 2015.